Theoretical Computer Science 113 (1993) 167–183
Elsevier

Probabilistic polynomials, AC^0 functions and the polynomial-time hierarchy*

Jun Tarui

Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK

Abstract

Tarui, J., Probabilistic polynomials, AC^0 functions and the polynomial-time hierarchy, Theoretical Computer Science 113 (1993) 167–183.

We show that, for every Boolean function $f(x_1, \ldots, x_n)$ in the class AC^0 and an arbitrary constant $k \geq 0$, there is a size-$O(n^{k+1})$ collection Ω of degree-$\log^{O(1)} n$ polynomials over $\mathbf{Z}$ in $x_1, \ldots, x_n$ such that, for each $x \in \{0,1\}^n$, when $p \in \Omega$ is randomly chosen, $f(x) = p(x)$ with probability at least $1 - 1/(3n^k)$, and, furthermore, if $f(x) = 0$ ($f(x) = 1$), then $p(x) = 1$ ($p(x) = 0$) with probability 0. Applying this result, we prove the following: (a) Every Boolean function in the class AC^0 can be computed with one-sided error at most $1/(3n^k)$ by some depth-two probabilistic circuits with a threshold gate at the root, $n^{\log^{O(1)} n}$ AND gates of fan-in $\log^{O(1)} n$ at the next level, and $(k+1)\log_2 n + O(1)$ random bits; it can also be computed, for an arbitrary constant $l \geq 0$, by some depth-three deterministic circuits with an OR gate at the root, at most $n/(\log_2 n)^l$ Threshold gates at the second level, and $n^{\log^{O(1)} n}$ AND gates of fan-in $\log^{O(1)} n$ at the third level. (b) For $\mathscr{C} = \mathrm{PP}$, $\mathrm{C}_=\mathrm{P}$, and $\mathrm{MOD}_m\mathrm{P}$, every language L in the polynomial-time hierarchy is $\mathscr{C}$-easy under a randomized many–one polynomial-time reduction; in fact, for $\mathscr{C} = \mathrm{PP}$ and $\mathrm{C}_=\mathrm{P}$, L is $\mathscr{C}$-easy under such a reduction with one-sided error.

1. Introduction and overview

Recently Toda [35] obtained the seminal result that $\mathrm{PH} \subseteq \mathrm{P}^{\#\mathrm{P}[1]}$, i.e., every language in the class PH, the polynomial-time hierarchy [21, 31], can be recognized by some polynomial-time Turing machine that makes one query to some function in the class # P [37]. He obtained this result by showing that $\mathrm{PH} \subseteq \mathrm{BP} \cdot \oplus \mathrm{P}$ and that

Correspondence to: J. Tarui, Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK. Email: jun@dcs.warwick.ac.uk.

* Supported in part by the ESPRIT II BRA Programme of the EC under contract # 7141 (ALCOM II). Part of the work was done while the author was a student at the University of Rochester, and was supported in part by NSF grant CDA-8822724. A preliminary version of this paper appeared as [33].

$BP\cdot\oplus P\subseteq P^{\#P[1]}$. (Since $P^{\#P[1]}\subseteq P^{PP}$, it also follows that $PH\subseteq P^{PP}$. Definitions of the operator BP, the classes $\oplus$ P ("Parity-P") and PP, and other classes mentioned in this section are given in Section 2.)

The result that $PH\subseteq BP\cdot\oplus P$ is interesting in its own right. It says that every language in PH is $\oplus$P-easy under a randomized many–one polynomial-time reduction: For every language L in PH, there is a polynomial $p(n)$ and a nondeterministic polynomial-time Turing machine M whose input is of the form (x, ρ), where x is a regular input string and ρ is a string chosen at random from $\{0, 1\}^{p(|x|)}$, such that, for every $x\in\{0, 1\}^*$, if $x\in L$ ($x\notin L$), then $\#M(x, \rho)$, the number of accepting paths of M on (x, ρ), is odd (even) with high probability. In this paper, we improve this result and show that "is odd/even" above can be replaced by "$=f(x)+1/f(x)$": For every language L in PH, there is a polynomial $p(n)$ and a machine $M(x, \rho)$ as above and a polynomial-time-computable integer-valued function $f(x)$ such that, for each $x\in\{0, 1\}^*$, $f(x)$ is always even and (1) if $x\in L$ ($x\notin L$), then $\#M(x, \rho)=f(x)+1(f(x))$ with high probability, and, furthermore, (2) if $x\in L$ ($x\notin L$), then $\#M(x, \rho)=f(x)$ ($f(x)+1$) with probability 0. It follows readily that, for $\mathscr{C}=$ PP, $C_=P$, and MOD_mP (m is arbitrary; $\oplus P=MOD_2P$), every language L in PH is $\mathscr{C}$-easy under a randomized many–one polynomial-time reduction; in fact, for $\mathscr{C}=$ PP and $C_=P$, L is $\mathscr{C}$-easy under such a reduction with one-sided error.

The well-studied class AC^0 consists of Boolean functions computable by constant-depth polynomial-size circuits with NOT gates and unbounded fan-in AND and OR gates. There is a well-known connection [14] between the class PH and the class AC^0 (more precisely, corresponding to PH is the class qAC^0 obtained by taking the size bound to be *quasi*polynomial, i.e, $n^{\log^{O(1)}n}$). After Toda showed that $PH\subseteq BP\cdot\oplus P$, Allender [1] showed that every function in AC^0 can be computed with small error by depth-two probabilistic circuits with a PARITY gate at the root, $n^{\log^{O(1)}n}$ AND gates of fan-in $\log^{O(1)}n$ at the next level, and $n^{O(1)}$ random bits; it can also be computed by depth-three deterministic threshold circuits with a Threshold gate at the root, $n^{\log^{O(1)}n}$ Threshold gates at the second level, and $n^{\log^{O(1)}n}$ AND gates of fan-in-$\log^{O(1)}n$ at the third level.

Allender [1] used the Razborov–Smolensky probabilistic simulation of AND and OR (cf. Remark 3.3). Later, using the simulation based on Valiant and Vazirani's lemma (in terms of "mod 2"; explained below), Allender and Hertrampf [2] and, independently, Kannan et al. [19] obtained uniform versions of Allender's results above. The works reported in this paper are independent of [2, 19]; we obtain results (implicit in the proof of Theorem 4.1) on nonuniform and uniform circuits that are stronger than the results in [1, 2, 19] in the same way that our main result on PH is stronger than the assertion that $PH\subseteq BP\cdot\oplus P$, as explained above. Consequently, we can show that every function in AC^0 can be computed with small, one-sided error by depth-two probabilistic circuits with a Threshold gate at the root, $n^{\log^{O(1)}n}$ AND gates of fan-in $\log^{O(1)}n$ at the next level, and $O(\log n)$ random bits; it can also be computed, for an arbitrary constant $l\geqslant 0$, by some depth-three deterministic circuits with an OR

gate at the root, at most $n/(\log_2 n)^l$ Threshold gates at the second level, and $n^{\log^{O(1)} n}$ AND gates of fan-in $\log^{O(1)} n$ at the third level.

The following is a rough description of main proof ideas used in this paper.

The key ingredient in Toda's proof that $\mathrm{PH} \subseteq \mathrm{BP} \cdot \oplus \mathrm{P}$ is a new application of Valiant and Vazirani's lemma [38] (restated as Lemma 3.1 in this paper). The conclusion of their lemma is of the form "$\ldots |S_i| = 1$ for some i", but Toda (and the authors of [2] and [19]) used only a weaker conclusion of the form "$\ldots |S_i| \equiv 1 \pmod 2$ for some i."

We can exploit the full strength of Valiant and Vazirani's lemma by interpreting it as showing that, over an *arbitrary* ring (not just over $\mathbf{Z}/2\mathbf{Z}$, the ring of integers mod 2), in particular over $\mathbf{Z}$, there is a *low-degree probabilistic polynomial* with *small sample space* that computes $\mathrm{OR}(x_1, \ldots, x_n)$ with high probability, i.e., there is a small collection Ω of degree-$\log^{O(1)} n$ polynomials in $x_1, \ldots, x_n$ such that, for every $x \in \{0, 1\}^n$, when $p \in \Omega$ is randomly chosen, $\mathrm{OR}(x_1, \ldots, x_n) = p(x_1, \ldots, x_n)$ with high probability.

Once we make this observation, it is straightforward to show that every function $f(x_1, \ldots, x_n)$ in AC^0 can be computed with small error by a low-degree probabilistic polynomial with small sample space. What we actually show is that, over $\mathbf{Z}$, there is a similar probabilistic polynomial p for f that has the following additional property, which we call *Boolean guarantee*: With probability 1, $p(x) \in \{0, 1\}$ implies that $f(x) = p(x)$, or, equivalently, if $f(x) = 0$ ($f(x) = 1$), then $p(x) = 1$ ($p(x) = 0$) with probability 0. To achieve Boolean guarantee, we construct a probabilistic polynomial in such a way that errors are controlled nicely under composition.

By applying this result concerning probabilistic polynomials, we obtain the results on polynomial-time hierarchy and the circuits mentioned above; in particular, Boolean guarantee yields property (2) stated in the second paragraph of this section and *one-sided* error in probabilistic threshold circuits.

The following works are independent of the work reported in this paper: In terms of the polynomial-time hierarchy, Toda and Ogiwara [36] have obtained similar results. In terms of circuits, building upon the work of Toda and Ogiwara above, Beigel et al. [6] have obtained similar results. In both cases, we obtain results that are somewhat stronger using the idea of achieving Boolean guarantee, which appears only in this work; otherwise, the proof ideas used are similar.

The work reported in this paper was partly motivated by, and extends, the work of Ogiwara [24], in which he showed that $\Pi_2^{\mathrm{P}} \subseteq \mathrm{BP} \cdot \mathrm{C}_{=}\mathrm{P}$ using an idea different from ours. (The polynomial-time hierarchy [21, 31] is the infinite union of Σ_i^{P} or its complement Π_i^{P} ($i = 0, 1, \ldots$). Definitions can also be found, e.g, in [18].)

The paper is organized as follows. After the next preliminary section, in Section 3 we show that every function in AC^0 can be computed with small error and with Boolean guarantee by low-degree probabilistic polynomials with small sample space. Consequences in terms of threshold circuits and the polynomial-time hierarchy are explained in Sections 4 and 5, respectively.

2. Preliminaries

Logarithms in the paper are to the base 2. Whenever we consider a sequence $\langle f_n \rangle_{n=0}^{\infty}$ of Boolean functions, a sequence $\langle C_n \rangle_{n=0}^{\infty}$ of circuits, or a sequence $\langle p_n \rangle_{n=0}^{\infty}$ of polynomials, f_n is a function from $\{0,1\}^n$ to $\{0,1\}$, C_n is a circuit for n Boolean variables, and p_n is a polynomial in n variables. We will simply write, e.g., "a family $\{f_n\}$ of Boolean functions" when we mean such a sequence of Boolean functions. We let $\mathbf{Z}$ and $\mathbf{Z}/m\mathbf{Z}$ denote, respectively, the ring of integers and the ring of integers mod m. The set of reals is denoted by $\mathbf{R}$. For a ring R, $R[x_1, \ldots, x_n]$ denotes the set of polynomials over R in $x_1, \ldots, x_n$. For a polynomial p over $\mathbf{Z}$, we define the *norm* of p to be the sum of the absolute values of p's coefficients.

We assume that the reader is familar with the standard terminology about circuits. For a gate G or a circuit C, we let G and C also denote the functions computed by G and C, respectively. As usual, we assume that, for a (probabilistic) circuit for n Boolean variables $x_1, \ldots, x_n$ (and v random bits $r_1, \ldots, r_v$), $2n$ positive and negated literals $x_1, \overline{x_1}, \ldots, x_n, \overline{x_n}$ (and $r_1, \overline{r_1}, \ldots, r_v, \overline{r_v}$) are given as inputs for a circuit, and NOT gates appear only as negated literals. We visualize a circuit as having the output gate *at the top*. The gates "immediately above" the inputs are *at the bottom*.

For a probabilistic circuit C for n variables and a Boolean function $f: \{0,1\}^n \to \{0,1\}$, *C computes f with error at most ε* if, for each $x \in \{0,1\}^n$, $f(x) = C(x)$ with probability at least $1-\varepsilon$; and C computes f with *one-sided* error at most ε if, in addition, for each $x \in \{0,1\}^n$, if $f(x)=0$, then $C(x)=0$ with probability 1. Similarly, C computes f with *reverse*-one-sided error at most ε if when $f(x)=1$, then $C(x)=1$ with probability 1.

For input Boolean variables $y_1, \ldots, y_l$, a Threshold$_t$ gate outputs 1 if $\sum y_i \geq t$, and 0 otherwise; an Exact$_k$ gate outputs 1 if $\sum y_i = k$, and 0 otherwise; and a MOD$_m$ gate outputs 1 if $\sum y_i \not\equiv 0 \pmod m$, and 0 otherwise.

The negation of OR is denoted by NOR. For convenience, instead of AND/OR circuits, we mainly deal with circuits that consist solely of NOR gates, using the following straightforward conversion of an AND/OR circuit in which AND and OR gates appear at alternating levels into a NOR circuit.

Fact 2.1. *Let C be an AND/OR circuit in which AND gates and OR gates appear at alternating levels. Let C' be the circuit obtained from C as follows: Replace every gate in C by a NOR gate except for this: If the bottom gates of C are AND gates, for each bottom gate G of the form* $\mathrm{AND}(u_1, \ldots, u_l)$, *where u_i's are input literals, replace G by a NOR gate of the form* $\mathrm{NOR}(\overline{u_1}, \ldots, \overline{u_l})$. *Then, for each gate G' in C' that has replaced a gate G in C, the following hold:*

$$G \text{ is an AND gate} \Rightarrow G'(x) = G(x),$$

$$G \text{ is an OR gate} \Rightarrow G'(x) = \overline{G(x)}.$$

In particular, the circuit C' computes either $C(x)$ or $\overline{C(x)}$.

Proof. By induction on $\delta = 1, \ldots, d$, we can easily show that the gates at depth $d-\delta$ satisfy the condition above. □

For a nondeterministic Turing machine M and an input x, we let $\# M(x)$ denote the number of accepting paths of M on x.

The "counting classes" PP, $C_=P$, MOD_mP, and $\oplus P$ (first considered, respectively, in [15], [39], [5, 10, 17], and [25]) are defined as follows. In what follows, M denotes a nondeterministic polynomial-time Turing machine and f denotes a polynomial-time-computable function from $\{0, 1\}^*$ to the set of nonnegative integers:

$$\mathrm{PP} = \{L: (\exists M, f)\ (\forall x) x \in L \Leftrightarrow \#M(x) \geqslant f(x)\},$$

$$\mathrm{C}_=\mathrm{P} = \{L: (\exists M, f)\ (\forall x) x \in L \Leftrightarrow \#M(x) = f(x)\},$$

$$\mathrm{MOD}_m\mathrm{P} = \{L: (\exists M)\ (\forall x) x \in L \Leftrightarrow \#M(x) \not\equiv 0 \pmod m\},$$

$$\oplus\mathrm{P} = \mathrm{MOD}_2\mathrm{P}.$$

For a predicate ϕ and a finite set S, let $\mathrm{Prob}[\phi(\rho): \rho \in_{\mathrm{R}} S]$ denote the probability that $\phi(\rho)$ is true when ρ is chosen uniformly at random from S.

The "probabilistic operators" BP, R, $\bar{\mathrm{R}}$, and ZP, which define, for any class $\mathscr{C}$ of languages, new classes of languages (written, e.g., $\mathrm{BP}\cdot\mathscr{C}$), are defined as follows. (Similar operators have been considered by Schöning [29], Zachos [40], and others. The definitions given below are consistent with the names of the standard classes BPP, RP and ZPP.)

Let $\mathscr{C}$ be a class of languages. The class $\mathrm{BP}\cdot\mathscr{C}$ consists of those languages L for which there exist a language $A \in \mathscr{C}$ and a polynomial $p(n)$ such that, for each $x \in \{0, 1\}^*$,

$$x \in L \Rightarrow \mathrm{Prob}[\langle x, \rho\rangle \in A: \rho \in_{\mathrm{R}} \{0, 1\}^{p(|x|)}] \geqslant 2/3;$$

$$x \notin L \Rightarrow \mathrm{Prob}[\langle x, \rho\rangle \in A: \rho \in_{\mathrm{R}} \{0, 1\}^{p(|x|)}] \leqslant 1/3.$$

Similarly, define the class $\mathrm{R}\cdot\mathscr{C}$ (and the class $\bar{\mathrm{R}}\cdot\mathscr{C}$) by the conditions

$$x \in L \Rightarrow \mathrm{Prob}[\langle x, \rho\rangle \in A: \rho \in_{\mathrm{R}} \{0, 1\}^{p(|x|)}] \geqslant 1/2 \quad (=1),$$

$$x \notin L \Rightarrow \mathrm{Prob}[\langle x, \rho\rangle \in A: \rho \in_{\mathrm{R}} \{0, 1\}^{p(|x|)}] = 0 \quad (\leqslant 1/2).$$

The operators R and $\bar{\mathrm{R}}$ correspond, respectively, to "one-sided error" and "reverse-one-sided error" defined for probabilistic circuits above (and defined for probabilistic polynomials in the next section). Finally, define the class $\mathrm{ZP}\cdot\mathscr{C}$ as

$$\mathrm{ZP}\cdot\mathscr{C} = \{L: L \in \mathrm{R}\cdot\mathscr{C} \text{ and } \bar{L} \in \mathrm{R}\cdot\mathscr{C}\}.$$

3. Probabilistic polynomials and AC^0 functions

Let R be a ring. A *probabilistic polynomial* $p(x_1, \ldots, x_n)$ over R is a random variable that is uniformly distributed over some finite multiset $\Omega = \{p_1, \ldots, p_s\}$, where

$p_i \in R[x_1, \ldots, x_n]$ $(1 \leqslant i \leqslant s)$. (A more general definition can be given, but for our purposes the definition above suffices.) The *degree* and the *norm* of a probabilistic polynomial p are, respectively, the maximum degree and the maximum norm of p_i $(1 \leqslant i \leqslant s)$. We call Ω the *sample space* of p and $s = |\Omega|$ the *sample size* of p. We sometimes describe a probabilistic polynomial p of sample size 2^v by specifying how a polynomial is sampled given v random bits, i.e., by giving a map from $\{0, 1\}^v$ to $R[x_1, \ldots, x_n]$; in this case, we speak of a probabilistic polynomial that *depends on v random bits.*

Let $p(x) = p(x_1, \ldots, x_n)$ be a probabilistic polynomial and let $f: \{0, 1\}^n \to \{0, 1\}$. We say that p *computes f with error at most ε* if, for each $x \in \{0, 1\}^n$, $f(x) = p(x)$ with probability at least $1 - \varepsilon$; p computes f with *one-sided* error at most ε if, in addition, for each $x \in \{0, 1\}^n$, if $f(x) = 0$, then $p(x) = 0$ with probability 1. Similarly, p computes f with *reverse*-one-sided error at most ε if when $f(x) = 1$, then $p(x) = 1$ with probability 1.

We say that $p(x)$ computes $f(x)$ *with Boolean guarantee* if, for each $x \in \{0, 1\}^n$, with probability 1, $p(x) \in \{0, 1\}$ implies that $f(x) = p(x)$, or, equivalently, if $f(x) = 0$ $(f(x) = 1)$, then $p(x) = 1$ $(p(x) = 0)$ with probability 0.

Probabilistic polynomials are over $\mathbf{Z}$ unless explicitly stated otherwise.

We will use the following lemma of Valiant and Vazirani [38, Theorem 2.4(i)] to prove our next lemma. Consider $\{0, 1\}^n$ to be an n-dimensional vector space over $\mathbf{Z}/2\mathbf{Z}$ in a natural way and let $\cdot$ denote the standard inner product over $\mathbf{Z}/2\mathbf{Z}$.

Lemma 3.1 (Valiant and Vazirani [38]). *Let S be an arbitrary nonempty subset of $\{0, 1\}^n$. Pick $w_1, \ldots, w_n$ independently and uniformly at random from $\{0, 1\}^n$ and, for each $i \in \{0, \ldots, n\}$, let*

$$S_i = \{v \in S: v \cdot w_1 = \cdots = v \cdot w_i = 0\} \quad (S_0 = S).$$

Then, with probability at least $1/4$, $|S_i| = 1$ for some i.

Lemma 3.2. *For an arbitrary $\varepsilon > 0$, there is a probabilistic polynomial $p(x_1, \ldots, x_n)$ of degree $O(\log(1/\varepsilon)\log n)$ and norm $n^{O(\log(1/\varepsilon)\log n)}$ that computes $\mathrm{NOR}(x_1, \ldots, x_n)$ with reverse-one-sided error at most ε and satisfies the following condition.*

Condition (*): *For any nonnegative integers $x_1, \ldots, x_n$, the following hold with probability 1:*

$$p(x_1, \ldots, x_n) \geqslant 0,$$

$$p(x_1, \ldots, x_n) = 1 \Rightarrow x_1 = \cdots = x_n = 0,$$

$$p(x_1, \ldots, x_n) = 0 \Rightarrow x_j = 1 \text{ for some } j \in \{1, \ldots, n\}.$$

Proof. Without loss of generality, assume that n is a power of 2. (If n is not a power of 2, let $N = 2^{\lceil \log n \rceil}$ and obtain $p'(x_1, \ldots, x_N)$, and then obtain $p(x_1, \ldots, x_n)$ from p' by setting $x_j = 0$ in p' for $j = n+1, \ldots, N$.)

We can identify the set $\{1, \ldots, n\}$ with the set $\{0, 1\}^{\log n}$ by, say, letting $j \in \{1, \ldots, n\}$ correspond to binary $(j-1)$, the standard $(\log n)$-bit binary representation of $j-1$. Thus, each $x \in \{0, 1\}^n$ can be thought of as the characteristic n-bit vector of some subset of $\{0, 1\}^{\log n}$; in particular, a point $x \in \{0, 1\}^n$ such that $x \neq (0, \ldots, 0)$ corresponds to a nonempty subset. Thus, we can apply Valiant and Vazirani's lemma as follows.

Pick $w_1, \ldots, w_{\log n}$ independently and uniformly at random from $\{0, 1\}^{\log n}$ and, for each $i \in \{0, \ldots, \log n\}$ and each $j \in \{1, \ldots, n\}$, let

$$f_j^{(i)} = \begin{cases} 1 & \text{if } \text{binary}(j-1) \cdot w_1 = \cdots = \text{binary}(j-1) \cdot w_i = 0 \ (r_j^{(0)} = 1), \\ 0 & \text{otherwise.} \end{cases}$$

Let

$$q(x_1, \ldots, x_n) = \prod_{i=0}^{\log n} (r_1^{(i)} x_1 + \cdots + r_n^{(i)} x_n - 1)^2.$$

By Valiant and Vazirani's lemma, for each nonzero $(x_1, \ldots, x_n) \in \{0, 1\}^n$, with probability at least 1/4, the following holds: For some $i \in \{0, \ldots, \log n\}$, there is a unique $j \in \{1, \ldots, n\}$ such that $x_j = 1$ and $r_j^{(i)} = 1$. Therefore, the probabilistic polynomial q computes NOR with reverse-one-sided error at most 3/4.

To satisfy condition (*) and to reduce the error bound to ε, take $l = O(\log(1/\varepsilon))$ large enough so that $(3/4)^l \leqslant \varepsilon$ and let

$$p(x_1, \ldots, x_n) = (x_1 + \cdots + x_n + 1) \prod_{k=1}^{l} q_k(x_1, \ldots, x_n),$$

where q_k's are l independent copies of q.

To check that p satisfies condition (*), let $x_1, \ldots, x_n$ be arbitrary nonnegative integers and let a $\{0, 1\}$-setting of $r_j^{(i)}$'s be arbitrary. Clearly, $p(x_1, \ldots, x_n)$ is nonnegative. If $p(x_1, \ldots, x_n) = 1$, then the first factor $(x_1 + \cdots + x_n + 1)$ must be one and, thus, $x_1 = \cdots = x_n = 0$. On the other hand, suppose that $p(x_1, \ldots, x_n) = 0$. Then, for some k, $q_k(x_1, \ldots, x_n) = 0$ and, thus, some factor of q_k is zero. From this it follows readily that $x_j = 1$ for some j.

Finally, one can easily check that the degree and the norm of p are $O(\log(1/\varepsilon) \log n)$ and $n^{O(\log(1/\varepsilon) \log n)}$, respectively. □

Remark 3.3. Razborov [28] and Smolensky [30] have observed that, over the field $\mathbf{Z}/p\mathbf{Z}$ (p prime), the probabilistic polynomial $q(x) = (r_1 x_1 + \cdots + r_n x_n)^{p-1}$ obtained by choosing $r_1, \ldots, r_n$ independently and uniformly at random from $\mathbf{Z}/p\mathbf{Z}$ computes $\text{OR}(x_1, \ldots, x_n)$ with one-sided error at most $1/p$. (Note that the sample space of q is large (of size p^n).) Smolensky [30] poses the question whether a similar construction is possible over a field with characteristic 0. The probabilistic polynomial p constructed in the proof above and $1-p$ compute NOR and OR, respectively, over an *arbitrary* ring; thus, we can answer the question in the affirmative.

Remark 3.4. The probabilistic polynomial p constructed in the proof above depends on $v=\Theta(\log(1/\varepsilon)\log^2 n)$ random bits and, thus, has sample size $(1/\varepsilon)n^{\Theta(\log n)}$. Since its construction is "explicit" and v is only of order $\log^{O(1)} n$ when $\varepsilon=\log^{-O(1)} n$, we can use p to obtain results in a "uniform" setting.

For results in the nonuniform setting, we can use a probabilistic polynomial with smaller sample space: By the following simple nonconstructive argument, we can reduce the sample size to $O(1/\varepsilon \cdot n)$ while maintaining the conditions on p. Further discussions of the degree and the sample size required for a (general or "explicit") probabilistic polynomial (over $\mathbf{Z}$ or $\mathbf{Z}/p\mathbf{Z}$) computing the OR function appear in [34].

Proposition 3.5. *Let $p(x_1, \ldots, x_n)$ be an arbitrary probabilistic polynomial computing $f: \{0,1\}^n \to \{0,1\}$ with error at most ε. Then, for any constant $\delta > 0$, there is a probabilistic polynomial p' of sample size $O(1/\varepsilon \cdot n)$ that computes f with error at most $(1+\delta)\varepsilon$.*

Proof. Fix $x \in \{0,1\}^n$ and consider $r_1(x), \ldots, r_N(x)$, where r_i's are N independent copies of p. Let E be the number of errors among the r_i's, i.e., let $E = |\{i: r_i(x) \neq f(x)\}|$. By the Chernoff bound ([12]; or see, e.g., [27, pp. 50–58]) on the tails of Bernoulli trials, we can take $N = O(1/\varepsilon \cdot n)$ and have

$$\text{Prob}(E > (1+\delta)\varepsilon N) < 2^{-n}.$$

Thus, there exists a multiset $\Omega = \{q_1, \ldots, q_N\}$ of ordinary polynomials such that, for *every* $x \in \{0,1\}^n$,

$$|\{i: q_i(x) \neq f(x)\}| \leqslant (1+\delta)\varepsilon N.$$

Clearly, Ω considered as a probabilistic polynomial yields the proposition. □

Note that, in the proof above, the sample space of p' is just a subset of the sample space of p and, thus, p' "inherits" properties of p. For example, the degree and the norm of p' do not exceed, respectively, the degree and the norm of p, and if error of p is (reverse-)one-sided, then error of p' is (reverse-)one-sided.

The following is the main result in this section.

Theorem 3.6. *For every family $\{f_n\}$ in AC^0 and an arbitrary fixed constant $k \geqslant 0$, there exists some family $\{p_n\}$ of probabilistic polynomials over $\mathbf{Z}$ such that p_n computes f_n with error at most $1/(3n^k)$ and with Boolean guarantee and that p_n has degree $\log^{O(1)} n$, norm $n^{\log^{O(1)} n}$ and sample size $O(n^{k+1})$.*

Proof. Let $f = \{f_n\}$ be in AC^0 and let $\{C_n\}$ be a family of constant-depth polynomial-size circuits consisting solely of NOR gates that computes f. Let $k \geqslant 0$. Fix n and let $d = \text{depth}(C_n) = O(1)$ and $N = \text{size}(C_n) = n^{O(1)}$.

Obtain, as in the proof of Lemma 3.2 (set $\varepsilon = 1/(4Nn^k) = n^{-O(1)}$), a probabilistic polynomial $q(y_1, \ldots, y_N)$ that has degree $O(\log^2 n)$ and norm $n^{O(\log^2 n)}$, depends on $v = O(\log^3 n)$ random bits $r_1, \ldots, r_v$, computes $\text{NOR}(y_1, \ldots, y_N)$ with error probability

at most $1/(4Nn^k)$, and satisfies condition (*) of Lemma 3.2. We can express the sample space of q as $\{q_r(y_1, \ldots, y_N): r \in \{0, 1\}^v\}$.

Pick $r \in \{0, 1\}^v$ at random and replace each NOR gate in C_n by q_r, starting from the bottom level as follows. Assume that we are working on a gate G of fan-in $L \leqslant N$ with wires (edges) coming from gates $G_1, \ldots, G_L$ and that gates $G_1, \ldots, G_L$ have been replaced by polynomials $g_1(x_1, \ldots, x_n), \ldots, g_L(x_1, \ldots, x_n)$. We replace G by the polynomial $g(x_1, \ldots, x_n)$ obtained from $q_r(y_1, \ldots, y_N)$ by substituting $g_1(x_1, \ldots, x_n), \ldots, g_L(x_1, \ldots, x_n)$ for $y_1, \ldots, y_L$, respectively, and setting $y_j = 0$ for $j = L+1, \ldots, N$.

In the end, all the NOR gates in C_n have been replaced by probabilistic polynomials that depend on the common random bits $r_1, \ldots, r_v$.

Let $p(x_1, \ldots, x_n)$ be the probabilistic polynomial that has replaced the output gate of C_n. One can easily check that p has degree $\log^{O(1)} n$ and norm $n^{\log^{O(1)} n}$. For each $x \in \{0, 1\}^n$, the probability that $p(x) \neq f_n(x)$ is bounded by the sum, over all the NOR gates in C_n, of the probability that q does not compute NOR correctly at a given NOR gate; thus, it is at most $1/(4n^k)$.

We show, by induction on $\delta = 1, \ldots, d$, that each probabilistic polynomial that has replaced a NOR gate G at depth $d - \delta$ computes $G(x)$ with Boolean guarantee. First note that, by condition (*) on q, it is clear that, for every $x \in \{0, 1\}^n$, each probabilistic polynomial evaluates to a nonnegative integer with probability 1. The base case $\delta = 1$ is immediate by condition (*). Assume that every probabilistic polynomial that has replaced a NOR gate G at depth $\geqslant d - \delta$ computes $G(x)$ with Boolean guarantee. Consider a NOR gate G at depth $d - (\delta + 1)$ with wires coming from gates $G_1, \ldots, G_L$, and let g and $g_1, \ldots, g_L$, respectively, be the probabilistic polynomials that have replaced those gates. Let $x \in \{0, 1\}^n$ and let a $\{0, 1\}$-setting of the random bits $r_1, \ldots, r_v$ be arbitrary. Suppose that $g(x) = 1$. Then, by condition (*) on q, $g_j(x) = 0$ for all $j \in \{1, \ldots, L\}$. By induction, $G_j(x) = 0$ for all j and, thus, $G(x) = \mathrm{NOR}(0, \ldots, 0) = 1$. On the other hand, suppose that $g(x) = 0$. Then, by condition (*) on q, there exists j such that $g_j(x) = 1$. By induction, $G_j(x) = 1$ and, since G is a NOR gate, $G(x) = 0$.

Finally, by Proposition 3.5 and the remarks made thereafter, by allowing an error bound to be $1/(3n^k)$ instead of $1/(4n^k)$, we can reduce the sample size to $O(n^{k+1})$ while maintaining the other conditions.

Remark 3.7. After seeing an earlier version of this paper, Beigel et al. [6, Appendix] suggested an alternative way to achieve Boolean guarantee that yields a simplification of the arguments used above.

Remark 3.8. We say that a (ordinary, not probabilistic) polynomial $p(x_1, \ldots, x_n)$ *approximates* $f: \{0, 1\}^n \to \{0, 1\}$ if $f(x) = p(x)$ except for a "small" fraction of points in $\{0, 1\}^n$. By a standard averaging argument, it follows from the theorem above that, for every $f \in \mathrm{AC}^0$ and an arbitrary ring R, there are low-degree polynomials over R that approximate f. Aspnes et al. [3, Section 5] have observed that the fact that such an approximation is possible over $\mathbf{R}$, together with their new results, yields an alternative

proof for the well-known result (see, e.g., [9]) that constant-depth AND/OR circuits require exponential size to compute PARITY.

We say that a polynomial $p(x_1, \ldots, x_n)$ over $\mathbf{R}$ *L_2-approximates* $f: \{0,1\}^n \to \{0,1\}$ if the L_2 norm of $f-p$ is "small". Linial et al. [20] have shown that every $f \in AC^0$ is L_2-approximable by some low-degree polynomials over $\mathbf{R}$. The probabilistic polynomial obtained in the proof of Theorem 3.6 may evaluate to a number whose absolute value is "large" when error occurs. Thus, the method therein does *not* yield an alternative proof for the L_2-approximation result of Linial et al. [20].

4. Threshold circuits and AC^0 functions

Theorem 4.1. *For every family $\{f_n\}$ in AC^0 and an arbitrary constant $k \geq 0$, there is a family $\{C_n\}$ of depth-two probabilistic circuits such that C_n computes f_n with one-sided error at most $1/(3n^k)$ and that C_n has a Threshold gate at the top, $n^{\log^{O(1)} n}$ AND gates of fan-in $\log^{O(1)} n$ at the bottom, and $(k+1)\log n + O(1)$ random bits.*

By converting probabilistic polynomials given in Theorem 3.6 into probabilistic circuits, we can easily obtain the theorem. First we explain one possible conversion. For a circuit C and an input y, let $\# C(y)$ denote the number of 1-valued wires coming into the output gate of C on input y.

Lemma 4.2. *Let $\{q_r(x_1, \ldots, x_n): r \in \{0,1\}^v\}$ be a collection of polynomials over $\mathbf{Z}$ of degree at most δ and norm at most N. Then there is a depth-two circuit C with input variables $x_1, \ldots, x_n, r_1, \ldots, r_v$ such that, for each $x \in \{0,1\}^n$ and each $r \in \{0,1\}^v$, $\# C(x, r) = q_r(x) + N$ and that C has, at the bottom level, at most $\delta \cdot N \cdot 2^v$ AND gates of fan-in at most δ.*

Proof. Let $r \in \{0,1\}^v$. Let L_r be the set of v literals of the form r_i or $\overline{r_i}$ whose conjunction is 1 only on r; for example, if $r = (1, 1, 0, \ldots, 0)$, then $L_r = \{r_1, r_2, \overline{r_3}, \ldots, \overline{r_v}\}$.

Let $q'_r(x) = q_r(x) + N$, and express $q'_r(x)$ as the sum of $\{0,1\}$-valued terms as illustrated below:

$$2x_1 - 2x_2x_3 + N = x_1 + x_1 + (1 - x_2x_3) + (1 - x_2x_3) + \overbrace{1 + \cdots + 1}^{N-2}.$$

For each term of the form $x_{i_1} \cdots x_{i_j}$, create an AND gate over the set $\{x_{i_1}, \ldots, x_{i_j}\} \cup L_r$. For each term of the form $(1 - x_{i_1} \cdots x_{i_j})$, create j AND gates over the sets

$$\{\overline{x_{i_1}}\} \cup L_r, \{x_{i_1}, \overline{x_{i_2}}\} \cup L_r, \ldots, \{x_{i_1}, \ldots, x_{i_{j-1}}, \overline{x_{i_j}}\} \cup L_r.$$

For each 1 term, create an AND gate over the set L_r.

Let S_r be the set of AND gates thus created, and let S be the union of S_r over $r \in \{0, 1\}^v$. Connect each AND gate in S to a single-output gate and obtain a depth-two circuit C. Clearly, C satisfies the conclusion of the lemma. □

Proof of Theorem 4.1. Let $\{f_n\}$ be in AC^0 and let $k \geqslant 0$. By Theorem 3.6, there is a family $\{p_n\}$ of probabilistic polynomials over $\mathbf{Z}$ such that p_n computes f_n with error at most $1/(3n^k)$ and with Boolean guarantee and that p_n has degree $\log^{O(1)} n$, norm $n^{\log^{O(1)} n}$ and sample size $O(n^{k+1})$.

Let $p'_n = -2p_n^2 + 3p_n$. Since the quadratic $s(y) = -2y^2 + 3y$ fixes 0 and 1 and $s(y) \leqslant -2$ for all integer $y \notin \{0, 1\}$, p'_n computes f_n with error at most $1/(3n^k)$ and with Boolean guarantee just like p_n, and, for each $x \in \{0, 1\}^n$, $p'_n(x) \leqslant 1$ with probability 1.

Express the sample space of p'_n as $\{q_r(x_1, \ldots, x_n): r \in \{0, 1\}^v\}$, where $v = (k+1)\log n + O(1)$. (We may assume that the sample size of p'_n is a power of 2.) Let $N = n^{\log^{O(1)} n}$ be the norm of p'_n. By Lemma 4.2, there is a depth-two circuit C such that, for each $x \in \{0, 1\}^n$ and each $r \in \{0, 1\}^v$, $\#C(x, r) = q_r(x) + N$. Regarding $r_1, \ldots, r_v$ as random bits and putting a Threshold_{N+1} gate as the output gate of C, we obtain a circuit satisfying the conclusion of the theorem. □

Remark 4.3. Using $-p'_n$ (or p_n as obtained in the proof of Theorem 3.6) instead of p'_n, we can obtain a circuit that computes f_n with *reverse*-one-sided error at most $\varepsilon = 1/(3n^k)$.

If an Exact gate or a MOD gate is used as the output gate, the following hold. In the proof above, if we put, as the output gate of C, (a) an Exact_{N+1} gate or (b) a MOD_m gate (together with one extra wire connecting it to constant 1 if $N + 1 \equiv 0 \pmod m$), then we obtain a circuit that computes f_n with error at most ε, where error is one-sided in case (a) and is in general *not* one-sided in case (b).

(1) For such a circuit with an Exact gate as the output gate, error can*not* be made reverse-one-sided in general, e.g., for $f_n = \text{OR}$.

(2) For such a circuit with a MOD_m gate as the output gate: for an arbitrary m, error can*not* be made one-sided in general, e.g., for $f_n = \text{AND}$. For a prime m, error can*not* be made reverse-one-sided in general, e.g., for $f_n = \text{OR}$. (For a composite m, at present we cannot prove a similar result; for more information, see [4].)

We can easily show (1) and (2) by converting such circuits to low-degree polynomials and using, e.g., the arguments in [32].

Remark 4.4. Let $\text{sgn}(x)$ denote the function from $\mathbf{R}$ to $\{0, 1\}$ defined as $\text{sgn}(x) = 1$ if $x \geqslant 0$, and 0 otherwise. Let $h_n(x_1, \ldots, x_n)$ be an OR of "disjoint" ANDs of size $n^{2/3}$. Minsky and Papert [22, "One-in-a-box" Theorem, p. 59] have shown that if $p(x_1, \ldots, x_n)$ is a polynomial over $\mathbf{R}$ such that, for each $x \in \{0, 1\}^n$, $\text{sgn}(p(x)) = h_n(x)$, then p must have *degree* $\Omega(n^{1/3})$. It follows that to compute h_n, a depth-two deterministic threshold circuit requires *bottom fan-in* $\Omega(n^{1/3})$ (even without any size bound). Since $\{h_n\}$ is clearly in AC^0 and, thus, by Theorem 4.1, is computable by depth-two

size-$n^{\log^{O(1)} n}$ *probabilistic* threshold circuits of bottom fan-in $\log^{O(1)} n$, such probabilistic circuits are strictly more powerful than their deterministic counterparts.

From depth-two probabilistic circuits, we can easily obtain depth-three deterministic circuits.

Proposition 4.5. *For every family* $\{f_n\}$ *in* AC^0 *and an arbitrary constant* $l \geqslant 0$, *there is a family* $\{D_n\}$ *of depth-three deterministic circuits such that* D_n *computes* f_n *and has an* OR *gate at the top, at most* $n/(\log n)^l$ *Threshold gates at the next level, and* $n^{\log^{O(1)} n}$ AND *gates of fan-in* $\log^{O(1)} n$ *at the bottom.*

Proof. Let $\{f_n\}$ be in AC^0 and let $l \geqslant 0$. By inspecting the proofs of Theorems 3.6 and 4.1, one can easily see that there is a family $\{C_n\}$ of depth-two probabilistic circuits such that C_n computes f_n with one-sided error at most $\varepsilon = 2^{-\log^l n}$ and that C_n has a Threshold gate at the top and $n^{\log^{O(1)} n}$ AND gates at the bottom.

Let $A = f_n^{-1}(1) \subseteq \{0, 1\}^n$. By a standard averaging argument, there exists a setting of the random bits in C_n such that the resulting deterministic circuit correctly outputs 1 except at most ε fraction of points in A. We can repeat this argument on the set of points in A that we "missed". Thus, we can obtain at most $n/(\log n)^l$ depth-two deterministic circuits whose OR computes f_n. □

The following proposition says that there is an AC^0 function such that depth-three deterministic threshold circuits with $n^{\log^{O(1)} n}$ gates of fan-in $\log^{O(1)} n$ at the bottom level must have $n^{\Omega(1)}$ Threshold gates at the top two levels to compute it.

Let $h_n(x_1, \ldots, x_n)$ be as in Remark 4.4.

Proposition 4.6. *For an arbitrary* $\varepsilon > 0$, *a depth-three deterministic threshold circuit computing* h_n *with* $n^{\log^{O(1)} n}$ *gates of fan-in* $\log^{O(1)} n$ *at the bottom level must have* $\omega(1) n^{1/3-\varepsilon}$ *Threshold gates at the top two levels.*

Proof. As explained in Remark 4.4, for a polynomial $p(x_1, \ldots, x_n)$ over $\mathbf{R}$, if $\operatorname{sgn}(p(x)) = h_n(x)$, then $\operatorname{degree}(p) = \Omega(n^{1/3})$.

Let D be a depth-three circuit as specified above that has only $c \cdot n^{1/3-\varepsilon}$ Threshold gates at the top two levels for a constant $c > 0$. Then, using Newman's [23] low-degree rational approximation for $\operatorname{sgn}(y)$ as used by Paturi and Saks [26] and by Beigel et al. [7], we can express $D(x)$ as $D(x) = \operatorname{sgn}(q(x))$ for some polynomial $q(x_1, \ldots, x_n)$ over $\mathbf{R}$ whose degree is $o(n^{1/3})$. Thus, D does not compute h_n. □

Remark 4.7. As mentioned in Remark 3.4, using the probabilistic polynomial constructed in the proof of Lemma 3.2, without applying the nonconstructive argument of Proposition 3.5, we can obtain "uniform" versions of Theorems 3.6 and 4.1 and Proposition 4.5.

In the uniform version of Theorem 3.6 (Theorem 4.1) obtained in this way, probabilistic polynomials (probabilistic circuits) depend on (use) $O(\log^3 n)$ random

bits; fixing these random bits in all the possible ways and taking the OR of the resulting $n^{O(\log^2 n)}$ depth-two threshold circuits yields a uniform version of Proposition 4.5.

5. Polynomial-time hierarchy

Theorem 5.1. *Let L be a language in* PH. *Then there exists a nondeterministic polynomial-time Turing machine M, a polynomial $q(n)$, and a polynomial-time-computable function $f: \{0,1\}^* \to \mathbf{Z}$ such that, for each $x \in \{0,1\}^*$, when $\rho \in \{0,1\}^{q(|x|)}$ is chosen uniformly at random,*

$$x \notin L \Rightarrow \#M(x,\rho) = \begin{cases} f(x) & \text{with probability at least } 2/3, \\ f(x)+1 & \text{with probability } 0, \end{cases}$$

$$x \in L \Rightarrow \#M(x,\rho) = \begin{cases} f(x)+1 & \text{with probability at least } 2/3, \\ f(x) & \text{with probability } 0. \end{cases}$$

First we explain some consequences.

The theorem, together with the trivial fact that the classes PH and PP are closed under complementation and the well-known, nearly trivial fact that $C_=P \subseteq PP$, yields the following corollary.

Corollary 5.2.

$$PH \subseteq R \cdot C_=P; \quad \textit{thus, } PH \subseteq ZP \cdot C_=P \subseteq ZP \cdot PP,$$

$$PH \subseteq (R \cdot PP \cap \bar{R} \cdot PP),$$

$$PH \subseteq BP \cdot MOD_m P \quad \textit{for every } m \geqslant 2.$$

Remark 5.3. The arguments used to obtain Corollary 5.2 from Theorem 5.1 are relativizable and, as will be clear below, so are the arguments used to obtain Theorem 5.1. (Thus, these results hold with respect to any oracle.) By the arguments explained in Remark 4.3, we can show a limitation of relativizable proof techniques for showing certain stronger assertions: We can show that there is an oracle with respect to which $PH \not\subseteq \bar{R} \cdot C_=P$, $PH \not\subseteq R \cdot MOD_m P$ for every m, and $PH \not\subseteq \bar{R} \cdot MOD_m P$ for prime m.

Another similar limitation is known in the following context. From the assertion that $PH \subseteq ZP \cdot C_=P$ in the corollary above, it follows immediately that $PH \subseteq ZPP^{C_=P}$. As mentioned in Section 1, Toda [35] showed (by relativizable techniques) that $PH \subseteq P^{PP}$. Tarui [32] and, independently, Green [16] considered whether the stronger assertion that $PH \subseteq P^{C_=P}$ (which implies both the above relations) holds, and showed that there is an oracle with respect to which the relation does *not* hold and, in particular, $BPP \not\subseteq P^{C_=P}$. (Note that $BPP \subseteq (\Sigma_2^P \cap \Pi_2^P)$.)

Remark 5.4. As mentioned in Section 1, Toda and Ogiwara [36] have obtained results similar to Corollary 5.2, except that for our results about $C_{=}P$ and PP are slightly stronger than their results that $PH \subseteq BP \cdot C_{=}P$ and $PH \subseteq BP \cdot PP$. They have also observed that $C_{=}P^{PH} \subseteq BP \cdot C_{=}P$, $PP^{PH} \subseteq BP \cdot PP$, and $MOD_m P^{PH} \subseteq BP \cdot MOD_m P$. These additional observations are due to them. We note here that the first two assertions can be strengthened by our techniques to the following form:

$$C_{=}P^{PH} \subseteq R \cdot C_{=}P, \qquad PP^{PH} \subseteq (R \cdot PP \cap \bar{R} \cdot PP), \qquad PP^{PH} \subseteq ZP \cdot PP.$$

Since the relations in Corollary 5.2 hold with respect to any oracle as mentioned above, a standard argument [8] yields the following corollary.

Corollary 5.5. *With respect to a random oracle R, the following hold with probability* 1:

$$PH^R \subseteq C_{=}P^R,$$

$$PH^R \subseteq PP^R,$$

$$PH^R \subseteq MOD_m P^R \quad \textit{for every } m \geqslant 2.$$

Now we turn to the proof of Theorem 5.1. First we explain the following simple process. A *formula* is a circuit in which every gate has fan-out 1. Let F be a depth-two arithmetic circuit with a $\times$-gate at the top and m $+$-gates of fan-in s at the bottom. A straightforward distribution of multiplication over additions, together with making copies of gates, yields an arithmetic formula F' with a $+$-gate at the top and s^m $\times$-gates of fan-in m at the bottom. Call this conversion *switch*.

Example 5.6. Switching $(x+y)(x+y)$ yields $x^2+xy+yx+y^2$.

Let F be a depth-$2d$ arithmetic circuit with $\times$-gates of fan-in m and $+$-gates of fan-in s, where $\times$-gates and $+$-gates appear at alternating levels starting with a $\times$-gate at the top. By repeatedly copying gates and applying switches at the top two levels, we can convert F to a depth-two formula F' with a $+$-gate at the top and $s^{m+m^2+\cdots+m^d}$ $\times$-gates of fan-in m^d at the bottom. Call this conversion of F into F' *expansion*.

Example 5.7. The formula F' obtained by expanding

$$F = [(\cdot+\cdot)(\cdot+\cdot)+(\cdot+\cdot)(\cdot+\cdot)][(\cdot+\cdot)(\cdot+\cdot)+(\cdot+\cdot)(\cdot+\cdot)]$$

has $2^6 = 2^{2+2^2}$ terms of degree 2^2.

In the above example, each of the 2^{2+2^2} terms in F' can be specified by $2+2^2$ bits in the following way. With the process of top–down expansion of F in mind, let the first bit specify which of the two terms in the left square-bracketed factor is to be "picked up", let the second bit do the same for the right square-bracketed factor, and let the next four bits specify similarly for the four factors of the form $(\cdot+\cdot)(\cdot+\cdot)(\cdot+\cdot)(\cdot+\cdot)$.

Proof of Theorem 5.1. Let L be a language in PH and let $\bar{M}$ be an alternating polynomial-time Turing machine [11] that accepts L with constant number of alternations. For an input x of length n, the computation tree of $\bar{M}$ on x can be viewed as a constant-depth AND/OR tree with $2^{\text{poly}(n)}$ leaves such that each leaf corresponds to one computation path and has value 1 or 0 according to whether that path leads to acceptance or rejection and $\bar{M}$ accepts x if and only if this AND–OR tree evaluates to 1. By a straightforward conversion of an AND/OR tree to an NOR tree (Fact 2.1), we may assume that there is a fixed integer d, a polynomial $p(n)$, and a machine M' such that the following hold. For each input x of length n, M' first guesses d binary strings $y_1, \ldots, y_d$, each of length $p(n)$, then performs a deterministic polynomial-time computation on $w = \langle x, y_1, \ldots, y_d \rangle$ and accepts or rejects w; if we define $2^{d \cdot p(n)}$ Boolean values $z_1(x), \ldots, z_{2^{d \cdot p(n)}}(x)$ according to acceptance/rejection on the $2^{d \cdot p(n)}$ computation paths, $x \in L$ if and only if the canonical depth-d complete $2^{p(n)}$-ary NOR tree C on $z_1(x), \ldots, z_{2^{d \cdot p(n)}}(x)$ evaluates to 1.

Consider replacing, as in the proof of Theorem 3.6, every NOR gate in C by a probabilistic polynomial $R(u) = R(u_1, \ldots, u_{2^{p(n)}})$ of the form

$$R(u_1, \ldots, u_{2^{p(n)}}) = (u_1 + \cdots + u_{2^{p(n)}} + 1) \prod_i (r_1^{(i)} u_1 + \cdots + r_{2^{p(n)}}^{(i)} u_{2^{p(n)}} - 1)^2$$

such that R has degree $\delta(n) = O(p^2(n))$, $r_j^{(i)}$'s are $\{0, 1\}$-valued random variables determined by $q(n) = O(p^3(n))$ random bits, and the resulting depth-$2d$ (probabilistic) arithmetic circuit F computes C with error at most $1/3$ and with Boolean guarantee. In F, $\times$-gates and $+$-gates have fan-in $\delta(n)$ and $2^{p(n)} + 1$, respectively. Let F' be the depth-two formula obtained by expanding F. As explained after Example 5.7, each term t in F' can be specified by at most

$$\begin{aligned} l &= (p(n)+1)\delta(n) + (p(n)+1)\delta^2(n) + \cdots + (p(n)+1)\delta^d(n) \\ &= O(p(n) \cdot \delta^d(n)) = O(p^{2d+1}(n)) \end{aligned}$$

bits and has the form

$$t = a \cdot r_{j_1}^{(i_1)} \cdots r_{j_\nu}^{(i_\nu)} z_{k_1}(x) \cdots z_{k_\mu}(x) \quad (a = 1 \text{ or } -1),$$

where $\nu, \mu \leqslant \delta^d(n) = O(p^{2d}(n))$.

Let M be a nondeterministic polynomial-time Turing machine that operates as follows. Given an input x of length n and a random string ρ of length $q(n)$ (i.e., $q(n)$ random bits), M guesses l bits and obtain a term t. Then M computes the Boolean value

$$b = r_{j_1}^{(i_1)} \cdots r_{j_\nu}^{(i_\nu)} z_{k_1}(x) \cdots z_{k_\mu}(x)$$

in polynomial time and acts as follows.

Case $a = 1$: M accepts if $b = 1$ and rejects if $b = 0$.

Case $a = -1$: M accepts if $b = 0$ and rejects if $b = 1$.

For each term t with $a = -1$, according to whether $t = 0$ or -1, M produces one or zero accepting path, respectively, thus "padding" the number of accepting paths by

1 as in Lemma 4.2. Let $T(n)$ be the number of terms in F' with leading coefficient -1. It is easy to see that $T(n)$ is computable in time polynomial in n. The machine M, together with the polynomial $q(n)$ and the function $f(x)=T(|x|)$, satisfies the conclusion of the theorem. □

Remark 5.8. Using the framework of "Gap P" functions introduced by Fenner et al. [13] (in particular, using closure properties 3 and 4 [13, pp. 32–33]), we can somewhat simplify the proof above.

Acknowledgment

I thank Richard Beigel, Noam Nisan, Mitsunori Ogiwara, Dan Spielman, and Seinosuku Toda for valuable discussions. I also thank Joel Seiferas and the anonymous referees for suggestions for improving the presentation of the paper.

References

[1] E. Allender, A note on the power of threshold circuits, in: *Proc. 30th Ann. IEEE Symp. on Foundations of Computer Science* (1989) 580–584; a journal version is to appear as: E. Allender and U. Hertrampf, Depth reductions for circuits of unbounded fan-in, *Inform. and Comput.*

[2] E. Allender and U. Hertrampf, On the power of uniform families of constant depth threshold circuits, in: *Proc. 15th Internat. Symp. on Mathematical Foundations of Computer Science*, Lecture Notes in Computer Science, Vol. 452 (Springer, Berlin, 1990) 158–164; a journal version is to appear as: E. Allender and U. Hertrampf, Depth reductions for circuits of unbounded fan-in, *Inform. and Comput.*

[3] J. Aspnes, R. Beigel, M. Furst and S. Rudich, The expressive power of voting polynomials, in: *Proc. 23rd Ann. ACM Symp. on the Theory of Computing* (1991) 402–409.

[4] D. Barrington, R. Beigel and S. Rudich, Representing Boolean functions as polynomials modulo composite numbers, in: *Proc. 24th Ann. ACM Symp. on the Theory of Computing* (1992) 455–461.

[5] R. Beigel and J. Gill, Counting classes: thresholds, parity, mods, and fewness, *Theoret. Comput. Sci.* **103** (1992) 3–23.

[6] R. Beigel, N. Reingold and D. Spielman, The perceptron strikes back, in: *Proc. 6th Ann. IEEE Conf. on Structure in Complexity Theory* (1991) 286–291.

[7] R. Beigel, N. Reingold and D. Spielman, PP is closed under intersection, in: *Proc. 23rd Ann. ACM Symp. on the Theory of Computing* (1991) 1–9.

[8] C. Bennett and J. Gill, Relative to a random oracle A, $P^A \neq NP^A \neq co\text{-}NP^A$ with probability 1, *SIAM J. Comput.* **10** (1981) 96–113.

[9] R. Boppana and M. Sipser, The complexity of finite functions, in: J. van Leeuwen, ed., *Handbook of Theoretical Computer Science, Vol. A: Algorithms and Complexity* (Elsevier, Amsterdam, 1990) 757–804.

[10] J. Cai and L. Hemachandra, On the power of parity polynomial time, *Math. Systems Theory* **23** (1990) 95–106.

[11] A. Chandra, D. Kozen and L. Stockmeyer, Alternation, *J. ACM* **28** (1981) 114–133.

[12] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Statistics* **23** (1952) 493–507.

[13] S. Fenner, L. Fortnow and S. Kurtz, Gap-definable counting classes, in: *Proc. 6th Ann. IEEE Conf. on Structure in Complexity Theory* (1991) 30–42.

[14] M. Furst, J. Saxe and M. Sipser, Parity, circuits and the polynomial-time hierarchy, *Math. Systems Theory* **17** (1984) 13–27.
[15] J. Gill, Computational complexity of probabilistic Turing machines, *SIAM J. Comput.* **6** (1977) 675–695.
[16] F. Green, On the power of deterministic reductions of $C_{=}P$, *Math. Systems Theory*, to appear.
[17] U. Hertrampf, Relation among MOD-classes, *Theoret. Comput. Sci.* **74** (1990) 325–328.
[18] D. Johnson, A catalog of complexity classes, in: J. van Leeuwen, ed., *Handbook of Theoretical Computer Science, Vol. A: Algorithms and Complexity* (Elsevier, Amsterdam, 1990) 67–161.
[19] R. Kannan, H. Venkateswaran, V. Vinay and A. Yao, A circuit-based proof of Toda's theorem, *Inform. and Comput.*, to appear.
[20] N. Linial, Y. Mansour and N. Nisan, Constant depth circuits, Fourier transform, and learnability, in: *Proc. 30th Ann. IEEE Symp. on Foundations of Computer Science* (1989) 574–579.
[21] A. Meyer and L. Stockmeyer, The equivalence problem for regular expressions with squaring requires exponential time, in: *Proc. 13th Ann. IEEE Symp. on Switching and Automata Theory* (1972) 125–129.
[22] M. Minsky and S. Papert, *Perceptrons* (MIT Press, Cambridge, MA, 1969).
[23] D. Newman, Rational approximation to $|x|$, *Michigan Math. J.* **11** (1964) 11–14.
[24] M. Ogiwara, On the computational power of exact counting, manuscript, 1990.
[25] C. Papadimitriou and S. Zachos, Two remarks on the power of counting, in: *Proc. 6th GI Conf. on Theoretical Computer Science*, Lecture Notes in Computer Science, Vol. 145 (Springer, Berlin, 1983) 269–276.
[26] R. Paturi and M. Saks, On threshold circuits for parity, in: *Proc. 31st Ann. IEEE Symp. on Foundations of Computer Science* (1990) 397–404.
[27] P. Raghavan, Lecture notes on randomized algorithms, Research Report RC 15340 (# 68237), IBM, 1990.
[28] A. Razborov, Lower bounds on the size of bounded depth networks over a complete basis with logical addition, *Mat. Zametki* **41** (1987) 598–607 (in Russian); an English translation appears in: *Math. Notes* **41** (1987) 333–338.
[29] U. Schöning, Probabilistic complexity classes and lowness, *J. Comput. System Sci.* **39** (1988) 84–100.
[30] R. Smolensky, Algebraic methods in the theory of lower bounds for Boolean circuit complexity, in: *Proc. 19th Ann. ACM Symp. on the Theory of Computing* (1987) 77–82.
[31] L. Stockmeyer, The polynomial-time hierarchy, *Theoret. Comput. Sci.* **3** (1977) 1–12.
[32] J. Tarui, Degree complexity of Boolean functions and its applications to relativized separations, in: *Proc. 6th Ann. IEEE Conf. on Structure in Complexity Theory* (1991) 382–390.
[33] J. Tarui, Randomized polynomials, threshold circuits, and the polynomial hierarchy, in: *Proc. 8th Ann. Symp. on Theoretical Aspects of Computer Science*, Lecture Notes in Computer Science, Vol. 480 (Springer, Berlin, 1991) 238–250.
[34] J. Tarui, On singleton selectors and probabilistic polynomials, manuscript, 1992.
[35] S. Toda, PP is as hard as the polynomial-time hierarchy, *SIAM J. Comput.* **20** (1991) 865–877; an earlier version appeared as: S. Toda, On the computational power of PP and $\oplus$ P, in: *Proc. 30th Ann. IEEE Symp. on Foundations of Computer Science* (1989) 514–519.
[36] S. Toda and M. Ogiwara, Counting classes are at least as hard as the polynomial-time hierarchy, *SIAM J. Comput.* **21** (1992) 316–328; an earlier version appeared in: *Proc. 6th Ann. IEEE Conf. on Structure in Complexity Theory* (1991) 2–12.
[37] L. Valiant, The complexity of computing the permanent, *Theoret. Comput. Sci.* **7** (1979) 189–201.
[38] L. Valiant and V. Vazirani, NP is as easy as detecting unique solutions, Theoret. Comput. Sci. **47** (1986) 85–93.
[39] K. Wagner, The complexity of combinatorial problems with succinct input representation, *Acta Inform.* **23** (1986) 325–356.
[40] S. Zachos, Probabilistic quantifiers, adversaries, and complexity classes: overview, in: *Structure in Complexity Theory*, Lecture Notes in Computer Science, Vol. 223 (Springer, Berlin, 1986) 383–398.